TRANSLATION NO. 2174

DATE: 25 March 1968

DEPARTMENT OF THE ARMY
Fort Detrick
Frederick, Maryland

UDC 577.3

THERMODYNAMICS OF DNA MOLECULES

Biofizika (Biophysics)
Vol. XII, No. 6, 1967 pages 1075-1080

I. P. Pavlotskiy

The purpose of this work is to find the thermodynamic functions of the DNA molecules. Thermodynamic function may be readily calculated if we know the statistical sum. In order to calculate the statistical sum, we should adopt a definite model.

1. We shall proceed from the Watson and Crick model(1). In this model, the DNA molecules have the form of two counter spirals linked by hydrogen bonds. The lengths of the hydrogen bonds are strictly fixed and, according to numerous experiments, are somewhat less than 3 Å. Let us denote the length of the hydrogen bond by r_0. The energy of the hydrogen bonds may be measured according to a spectrogram. In fact, when the bond is broken, there radiates a wave whose length is unambiguously linked with the bond energy. The bond energy depends, in turn, on the bond length. The spectrograms obtained during the breaking of the hydrogen bonds of DNA molecules have one clear line without any blurring. This means that the potential bond energy is fixed very strictly and the bond potential has the form of a narrow slit: it is equal to infinity everywhere, except when the bond length is equal to r_0. At this length $\Phi(r_0) = W_0 = \text{const}$.

The molecule spirals consist of successively connected links of four types. The differences in the types of the links are due to the presence of four different radicals: adenine, thymine, isanine and cytosine. The links of the spirals of an equilibrium molecule perform thermal vibrations. However, the nature of the interaction potential between the bonds imposes substantial limitations on this movement: as any organic polymers, the DNA has fixed valence angles, i.e., angles between the neighboring links in the planes passing through these links. Thus, pairs of neighboring links represent something like rigid triangles, and thermal vibrations are nothing

but the changes of the angles between such triangles. In more popular terms, it is customary to say that the (k+1)-th link rotates along a cone in relation to the k-th link, and angle φ_k at the cone is called the rotation angle about the k-th bond(2). The link lengths are assumed to be constant (not necessarily identical). It is evident that the constancy of the link lengths and valence angles leaves a single degree of freedom (angles φ_k) and the system is, essentially, a one-dimensional system, although the molecule is situated in the usual three-dimensional space. This fact is utilized in calculating the thermodynamic functions of the DNA molecules. For example, in Zimm's works(3) the discussion is reduced to the one-dimensional Ising's model.

Let us also note that a qualitative explanation of the spiral torsion of the chains consisting of a sequence of links may only be given in a completely general case(4). In this case, the potentials $\Phi(\varphi_k)$ of interaction between the neighboring triangles may be symmetrical and it is only necessary that there should be a certain force $f \neq 0$ applied to the ends of the chain. If the tensile force disappears, then the spirals may change into a ball; however this is prevented by the presence of friction. Strictly speaking, the spiral-like structure at $f \neq 0$ is a result of the correlations between the rotations of the links and represents a collective effect.

2. We shall also consider only the interaction between the neighboring triangles as essential. Generally speaking, the form of the interaction potentials is not known at the present time; however, there is no basis to assume that the interaction is of a long-range nature and, therefore, the approaching of the nearest neighbors should satisfy all the reasonable requirements. Moreover, let us suppose, as it is usually done(2,5), that Hamilton's function of the system may be represented in the form of a sum of two addends one of which depends only on the coordinates and the other only on the impulses.

Strict examination indicates that in the polymeric chain Hamiltonian, there is a third addend, where the impulses and the coordinates are mixed. This addend may be ignored only when the interaction potentials of the "triangles" have a shape of sufficiently narrow and deep wells(6). Naturally, this requirement is more strict than the approaching of the nearest neighbors, however, it is still a generally accepted requirement and we shall leave it in force. Thus, we assume that Hamilton's function of the DNA molecule may be written in the form of

$$H = T(p) + U(q). \qquad (1)$$

where p is an aggregate of the generalized impulses of the links, and q is an aggregate of their generalized coordinates; then the statistical sum has the form

$$\int_{(\Gamma)} e^{-\beta[T(p)+V(q)]}\, dp\, dq, \qquad (2)$$

where integration takes place over the entire phase space Γ and $\beta = 1/kT$ -- inverse temperature.

Integration over the impulses with the assumptions that have been made is done in an elementary manner(2), therefore, we shall consider the configuration part of the statistical sum or, as they say, the configuration integral

$$Q = \int_{(\Omega)} e^{-\beta U(q)}\, dq, \qquad (3)$$

where integration is performed over the entire configuration space Ω.

Let $\varphi_{k-1,k,k+1} \equiv \varphi_k$ be the solid angle between the rigid triangles, the first of which is constructed on the (k+1)-th and k-th links, and the second on the k-th and the (k-1)-th links. Then

$$Q = \int_0^{2\pi} \dots \int_0^{2\pi} d\varphi_1 \dots d\varphi_N d\varphi_1' \dots d\varphi_N' \exp\{-\beta[U_1(\varphi_1, \dots, \varphi_N) + U_2(\varphi_1', \dots, \varphi_N') + W]\} \qquad (3a)$$

Here, the unaccented φ_i belong to the same spiral, and the accented φ_i' -- to the other spiral; $N_1 + 1$ and $N_2 + 1$ are the numbers of links of the first and the second spirals, respectively, and W -- is a potential interaction energy of the spirals.

3. Let us make a few remarks regarding W. If the DNA molecule is isolated and the number of the hydrogen bonds of the linking spirals is equal to k, then on the basis of the remarks in Paragraph 1.

$$W = kw_0.$$

The situation is different if the molecule is placed in a solution. Following the example of Frank-Kamenetskiy(7), let us call a tie any factor that changes the difference of free energies of the bound and free states of the nucleotide pair. If there are ties for the part of the configuration

integral connected with the integrand factor $\exp(-\beta W)$, we obtain an expression(7):

$$Z(n, N, \Delta, F_1, F_2) \simeq \lambda_0^{\frac{nN}{n+1}} \left\{ \frac{1}{(\lambda_1 - \lambda_0)} [\lambda_0 (\lambda_1 - 1) - ps(\lambda_0 - 1 + \sigma)] \right\}^{N/(n+1)} \qquad (4)$$

where $N = N_1 + 1 = N_2 + 1$, $n = N/k$, $p = \exp(\beta\Delta)$, $s = \exp(\beta F_1)$, $\sigma = \exp(-\beta F_2)$; F_1 is the difference of the free energies of the uncoupled and coupled pairs, if it is followed by a coupled pair, $F_1 + \Delta$ is the same for a pair with a tie, F_2 is the value subtracted from F_1 if the coupled pair is followed by an uncoupled pair and, finally, λ_0 and λ_1 are the characteristic numbers of the matrix

$$\begin{vmatrix} 1 & 1 \\ \sigma s & s \end{vmatrix}.$$

Expression (4) was obtained with certain assumptions formulated in work(7). In the same work a more general case is considered when the ties may move from one pair of bases to another.

We shall ignore the mechanism fastening the spirals, but will concentrate our attention on the role of thermal motion of the links making up the spiral. Moreover, keeping in mind the rigid nature of the hydrogen bonds between the spirals (see Paragraph 1), let us consider that these mechanisms are not bound. Then (3a) will assume the following form:

$$Q = Z(n, N, \Delta, F_1, F_2) \int_0^{2\pi} \dots \int_0^{2\pi} d\varphi_1 \dots d\varphi_N \cdot d\varphi_1' \dots \times$$
$$\times d\varphi_N' \exp\left(-\beta [U_1(\varphi_1, \dots, \varphi_N) + U_2(\varphi_1', \dots, \varphi_N')]\right) \qquad (3b)$$

We have assumed here that $N = N_1 + 1 = N_2 + 1$, and thus, integration from 0 to 2π is of a 2 N-multiple nature.

4. All that has been said above may be summed up in the following way. The configuration integral of the DNA molecule has the form of (3b) if the following assumptions are true: a) Hamilton's function may be presented in the form of (1) which is a generally accepted assumption; b) the correlation between the forces coupling the spirals and the thermal motion of the spirals' links is small.

All further calculations are of a completely precise nature.

In order to make this problem, formally, a unidimensional one, let us make the following substitution of the variables in the integrand; we assume

$$x_k = \sum_{1 \le i \le k} \Delta x_{i+1,\, i}, \tag{5}$$

where

$$\Delta x_{i+1,i} = 2l\cos\omega \frac{\sqrt{1 - \sin^2\omega \cdot \cos^2\varphi_i/2}}{\sin^2\omega + \cos\omega\sqrt{1 + 4\sin^2\omega/2}}, \tag{5a}$$

l is the distance between the geometrical centers of the i-th and the (i+1)-th triangles, and w is the value of the angle supplementing the valent one to π. Formula (5) is obtained from simple geometrical considerations (for more details see work(8).

It is seen from formula (5a) that the values of Δx are located between

$$x = \Delta x_{min} = 2l\cos^2\omega/(\sin^2\omega + \cos\omega\sqrt{1 + 4\sin^2\omega/2})$$

and

$$\Delta x_{max} = x/\cos\omega.$$

On the strength of the correspondence (5), we can proceed from the potentials of the interaction between the triangles $\Phi(\varphi)$ to the potentials of $\Phi(\Delta x)$. Limitations from top and bottom by the values Δx_k result in the requirement

$$\Phi(r) = \begin{cases} \infty & \text{at } r < x \\ \Phi(r) & \text{at } x \le r \le x/\cos\omega \\ \infty & \text{at } r > x/\cos\omega. \end{cases}$$

where $\Phi(x)$ is a function which is obtained from $\Phi(\varphi)$ when φ is substituted by x according in formulas (5).

Expression (3,b), after the above substitution, transforms as follows:

$$Q = 2\cdot 2^N \int_0^{l_1} \cdots \int_0^{l_N} d\varphi_1 \cdots d\varphi_N e^{-\beta \sum_{k=0}^{N} \Phi_k(x_k - x_{k-1})} \cdot D\left(\frac{\varphi_1, \ldots, \varphi_N}{\Delta x_1, \ldots, \Delta x_N}\right) \times$$
$$\times 2^N \int_0^{l_1} \cdots \int_0^{l_N} d\varphi'_1 \cdots d\varphi'_N \times e^{-\beta \sum_{k=0}^{N} \Phi_k(x_k - x_{k-1})} \cdot D\left(\frac{\varphi'_1, \ldots, \varphi'_N}{\Delta x_1, \ldots, \Delta x_N}\right) \tag{3c}$$

Factor 2^N in formula (3c) has appeared because of the ambiguity of the correspondence between φ and Δx. In fact, two angles φ differing by 180° correspond to one value of the distance Δx between the centers of the neighboring triangles, and, therefore, 2^N different methods of the distribution of the chain in space correspond to one definition of the coordinates (x_1, x_2, ..., x_N) at the same value of total potential energy.

The calculation of the jacobian of the transition from coordinates (φ_1, φ_2, ...) to coordinates (x_1, x_2, ...) is accomplished in an elementary way, inasmuch as, according to formulas (5) and (5a), only the elements on the main diagonal and under the main diagonal are nonzero.

As a result of this, we obtain

$$D\left(\frac{\varphi_1, \ldots, \varphi_N}{x_1, \ldots, x_N}\right) = \prod_{(i)} \frac{|x_{i+1} - x_i|}{\sqrt{a^2 - (x_{i+1} - x_i)^2}\sqrt{(x_{i+1} - x_i)^2 - a^2\cos^2\omega}}$$

$$\varepsilon = \frac{2|\cos\omega}{\sin^2\omega + \cos\sqrt{1 + 4\sin^2\omega/2}} \qquad (6)$$

Let us substitute (6) in (3c) introducing, for brevity, the following expression:

$$G_k(x_k - x_{k-1}) = \frac{2|x_k - x_{k-1}|\exp(-\beta\bar{U}_k(x_k - x_{k-1}))}{\sqrt{|a^2 - (x_k - x_{k-1})^2|\,|(x_k - x_{k-1})^2 - a^2\cos^2\omega|}}.$$

Then (3c) will assume the form of:

$$Q = \int_0^L dx_N \ldots \int_0^{x_2} dx_2 \int_0^{x_2} dx_1 \prod_{k=1}^{N} G_k(x_k - x_{k-1}) \times$$

$$\times \int_0^L dx'_N \ldots \int_0^{x'_2} dx'_2 \int_0^{x'_2} dx'_1 \prod_{k=1}^{N} G_k(x'_k - x'_{k-1}) \qquad (3d)$$

5. Inasmuch as integration in (3d) by the unaccented and the accented coordinates is done separately, it is sufficient to calculate the integral

$$q = \int_0^L dx_N \ldots \int_0^{x_2} dx_1 \prod G_k(x'_k - x_{k-1}).$$

by one spiral.

Let us take advantage of the method used earlier in work(9). Let us denote the Laplacian forms of the function by a tilde

$$\tilde{\varphi}(\xi) = \int_0^\infty e^{-\xi x} \varphi(x)\, dx, \qquad \mathrm{Re}(\xi) > 0$$

Let us introduce a sequence of functions

$$F_1 = G_1;\ F_M(x) = \int_0^x G_M(\tau)\, F_{M-1}(x-\tau)\, d\tau,$$

where each F_M is obtained as a convolution of F_{M-1} with G_M. Using the convolution theorem (the Laplacian form of convolution of two functions is equal to the product of their Laplacian forms), let us perform N times in sequence the Laplace transform of the expression of q. As a result, we shall obtain

$$\tilde{q}(\xi) = \prod_{i=1}^{N} \tilde{\sigma}_i(\xi),$$

and now the calculation of q is reduced to the integral

$$q = \frac{1}{2\pi i} \oint e^{\xi x} \prod_{i=1}^{N} \tilde{\sigma}_i(\xi)\, d\xi. \qquad (7)$$

Expression (7) permits further simplification, since only four types of links can occur in the DNA molecules (adenine, thymine, isanine, and cytosine). Inasmuch as potentials Φ_k depend on the mutual distribution of two neighboring triangles or, in other words, on three consecutive links, the number of potentials of various nature is determined by the number of combinations of four elements in groups of three differing either in the elements themselves or in their order with an additional condition: when the direct order of one combination coincides with the reverse order of the other, the combinations are considered to be identical. Let us explain this by an example: sequence (adenine, thymine, isanine) results in a different potential than sequence (thymine, adenine, isanine), however the potential corresponding to it is the same as for sequence (isanine, thymine, adenine).

A simple calculation results in a conclusion that the number of various combinations of the above-mentioned type is equal to 80. Thus, the number of various functions of σ_K is equal to 20.

If the code of the molecule is known, then it is known what portion of the γ_i groups of threes of the links belongs to each of the possible types.

It is evident that

$$\gamma_1+\gamma_2+\ldots+\gamma_{20}=1.$$

Thus, expression (7) may be written as

$$q=\frac{1}{2\pi i}\oint e^{\lambda z}[R(\xi)]^N\,d\xi. \tag{8}$$

where $R(\xi)=[\sigma_1(\xi)]^{\gamma_1 N}[\sigma_2(\xi)]^{\gamma_2 N}\ldots[\sigma_{20}(\xi)]^{\gamma_{20}N}$

$$\sigma_k(\xi)=\int_0^{\pi/\cos\alpha_k}\frac{2r\cos\left(-[\xi r+\beta\theta_k(r)]\right)}{\sqrt{a^2-r^2}\sqrt{r^2-a^2\cos^2\alpha}}\,dr \tag{8a}$$

In order to calculate the integral of (8) in asymptotic limits

$$L\to\infty,\ N\to\infty,\ \lim\frac{L}{N}=l$$

let us use the saddle point method.

We shall omit the calculations as they exactly repeat the scheme explained in works(8,9).

Let us give the final result

$$q(L)=e^{\beta fL}[R(\beta)]^N\cdot O\left(\frac{1}{\sqrt{N}}\right),$$

where f is force applied to the ends of the chain, β is inverse temperature, N is the number of links, and R is determined by formulas (8) and (8a).

If various L_1, N_1 and L_2, N_2 correspond to different spirals, then the expression for Q assumes the following form:

$$Q=\sum(n,N_1,N_2,\ldots,F_1,F_2)\,e^{\beta f(L_1+L_2)}[R(\beta)]^{N_1+N_2}.$$

Values L_1 and L_2, which are equal to

$$L_1=l_1N_1,\quad L_2=l_2N_2,$$

may be found from the "equation of state"(8):

$$I_{1,2}\cdot\beta+\frac{d}{d\beta}\ln R_{1,2}(\beta)=0. \qquad (9)$$

Finally,

$$Q=Z\left[d^{3N}R(\beta)\right]^{N_1+N_2}\cdot 0\left(\frac{1}{\sqrt{N_1}}\right)\cdot 0\left(\frac{1}{\sqrt{N_2}}\right), \qquad (10)$$

where $0\left(\frac{1}{\sqrt{N}}\right)$ are the factors which are unessential during further transition to lnQ.

In order to obtain the total statistical sum $\mathcal{E}$, it is necessary to multiply Q further by the part of the statistical sum connected with space integration of the impulses. If I_1, I_2, I_3, and I_4 are moments of inertia, and r_1, r_2, r_3, and r_4 are the radii of inertia of the links belonging to four different types, then the impulse part of the statistical sum has the form:

$$\int=\left[\frac{\left(I_1/r_1^2\right)^{\tau_1}\left(I_2/r_2^2\right)^{\tau_2}\left(I_3/r_3^2\right)^{\tau_3}\left(I_4/r_4^2\right)^{\tau_4}}{2\pi\beta\hbar^2}\right]^N\times\frac{(2N\tau_1)!\,(2N\tau_2)!\,(2N\tau_3)!\,(2N\tau_4)!}{2N!}, \qquad (11)$$

where h is the Planck constant and τ_i are the portions of links of each kind in the spiral:

$$\tau_1+\tau_2+\tau_3+\tau_4=1.$$

Thus, now we can write an expression for the statistical sum of the molecule (let $N_1=N_2$)

$$\mathcal{Z}=J\cdot Z\left[d^{3N}R(\beta)\right]^{2N}\cdot 0\left(\frac{1}{\sqrt{N}}\right), \qquad (12)$$

where expressions for J and Z are given by formulas (11) and (4).

Thermodynamic functions may be calculated by the standard formulas of statistical physics.

Free energy:

$$\Psi=-\beta^{-1}\ln\mathcal{Z}.$$

Entropy

$$S = A \ln Z - A\beta \frac{\partial \ln Z}{\partial \beta}$$

etc.

Summing up the above, we see that thermodynamic functions of the DNA molecule in a solution may be calculated if the assumptions briefly formulated in the beginning of paragraph 4 are true. For concrete calculations it is necessary to have explicit expressions for the interaction potentials of the links' groups of three $\theta_k(x)$, to know the codes of the molecules, i.e., numbers γ_k and τ_i. Formally, from the thermodynamic point of view, molecules can only differ in their different sets of numbers γ_k ($k = 1, 2, \ldots, 20$). The calculations themselves are reduced to very simple quadratic formulas (8a) which are considerably simpler than the problem of finding the fundamental functions and fundamental values of integro-differential equations in a method using the Markov's processes. Moreover, methods connected with Markov's chains require the assumption of the spirals' homogeneity(5), i.e., they do not take into consideration their coded nature and cannot "discriminate" the molecules of different nature.

* * *

In conclusion, I express my gratitude to S. V. Tyablikov, V. I. Ivanov and the late M. L. Tsetlin for their valuable discussions.

Bibliography

1. Crick, F., Watson, J. Proc. Roy. Soc., A 233, 80, 1954.

2. Vol'kenshtein, M. V., Konfiguratsionnaya statistika polimernykh tsepei (Configurational Statistics of Polymer Chains). Academy of Sciences Press, USSR, Moscow, 1959.

3. Zimm, B., J. Chem. Phys., 33, 1349, 1960.

4. Almazov, A. B., Pavlotsky, I. P. Ann. der Phys., 7, Folge. Bd. 17, Heft 5-6, 258, 1966.

5. Brishteyn, T. M., Ptitsyn, O. B. Konformatsiya molekul (Conformation of Molecules), "Mir" (World), Moscow, 1964.

6. Pavlotskiy, I. P. Statisticheskaya mekhanika polimernykh tsepei (Statistical Mechanics of Polymer Chains). Report at the International Congress of Mathematicians, Section 11, Moscow, 1966.

7. Frank-Kamenetskiy, M. D. Doklady (Reports) of the Academy of Sciences, USSR, 167, 1, 187, 1964.

8. Baranov, B. N., and Pavlotskiy, I. P. O konfiguratsionnoy statistike vysokomolekulyarnykh tsepey (On Configuration Statistics of High-Molecular Chains), Doklady, Academy of Sciences, USSR, 167, 5, 1964.

9. Pavlotskiy, I. P. Doklady, Academy of Sciences, USSR, 161, 66, 1965.

Received by Editors
10 XII 1965

I. P. Pavlotskiy, Department of Applied Mathematics, Institute of Mathematics imeni V. A. Steklov, Academy of Sciences, USSR.